



# A model for spectroscopic food sample analysis using data sonification

Hsein Kew<sup>1</sup>

Received: 18 July 2020 / Accepted: 31 December 2020 / Published online: 13 January 2021  
© The Author(s) 2021

## Abstract

In this paper, we propose a method to generate an audio output based on spectroscopy data in order to discriminate two classes of data, based on the features of our spectral dataset. To do this, we first perform spectral pre-processing, and then extract features, followed by machine learning, for dimensionality reduction. The features are then mapped to the parameters of a sound synthesiser, as part of the audio processing, so as to generate audio samples in order to compute statistical results and identify important descriptors for the classification of the dataset. To optimise the process, we compare Amplitude Modulation (AM) and Frequency Modulation (FM) synthesis, as applied to two real-life datasets to evaluate the performance of sonification as a method for discriminating data. FM synthesis provides a higher subjective classification accuracy as compared with to AM synthesis. We then further compare the dimensionality reduction method of Principal Component Analysis (PCA) and Linear Discriminant Analysis in order to optimise our sonification algorithm. The results of classification accuracy using FM synthesis as the sound synthesiser and PCA as the dimensionality reduction method yields a mean classification accuracies of 93.81% and 88.57% for the *coffee dataset* and the *fruit puree dataset* respectively, and indicate that this spectroscopic analysis model is able to provide relevant information on the spectral data, and most importantly, is able to discriminate accurately between the two spectra and thus provides a complementary tool to supplement current methods.

**Keywords** Spectroscopic analysis · Machine learning · Dimensionality reduction · Sound synthesiser · Audio processing

## 1 Introduction and literature survey

Sonification of spectral data has been explored in a number of research projects. Sturm for example presents work which explores patterns in spectral oceanography data Sturm (2002). The sonification of ocean buoy spectral data is able to provide important aspects of physical oceanography. De Campo et al. have used the sonification of quantum spectra in order to develop to a tool for classifying and describing baryon properties in the context of particle theory de Campo et al. (2005). Similarly, Vicinanza et al. Vicinanza et al. (2014) apply sonification to the analysis of stem cells.

The human ear has the capability to detect audio patterns and to recognise timbres. This will allow a further opportunity of introducing variables into the sound so that a listener can discriminate samples by listening to the audio clips, which represent information and data. There are many types of

sound synthesis methods which are used in auditory display design for direct parameter mapping. Kleiman-Wainer et al. developed a novel approach, using formant based vowel synthesis, in order to sonify golf-swing movements and improve the action of a golfer's swing Kleiman-Weiner and Berger (2006). Cassidy et al. also used the same synthesis method of formant based vowel synthesis in their approach to sonifying hyperspectral colon tissue Cassidy et al. (2004a) Cassidy et al. (2004b). They suggested there is potential for using vocal-like sounds for sonification where humans have the ability to easily identify such types of sounds. Zwicker claims that humans have the ability to recognize multiple simultaneously sounded vowels Zwicker (1984). Childs et al. applied frequency modulation synthesis to sonify hail storms, for the purpose of observing movements of hail storms and any changes in seasonal variation Childs and Pulkki (2003). One particular example of their sonification model of mapping, is that the size of the hail was mapped to the duration and loudness of the FM format instrument with the use of parameter-driven FM synthesis. The same FM synthesis method, was used in sonifying optical coherence tomography data and images of human tissue, for the purpose of discriminating between human

✉ Hsein Kew  
Hsein.Kew@bcu.ac.uk

<sup>1</sup> Digital Media Technology Lab, Birmingham City University, Millennium Point, Birmingham B4 7XG, UK

adipose and tumour tissues Ahmad et al. (2010). In addition, O'Neill applied both AM synthesis and FM synthesis as the synthesis methods for interactive sonification of 2-dimensional images, in order to gain spatial awareness of the image structure O'Neill and Ng (2008).

Spectroscopic analysis has been applied to a diverse range of such as monitoring long-term food storage and on shelf life Beghi et al. (2014), estimating the ripeness of different fruit Giovenzana et al. (2014), verification of the origin of raw materials and finished products Oliveri et al. (2011) and inspection of food quality Uslu et al. (2016). In all of these tasks, a suitable classification strategy using multivariate analysis *MVA* for analytical data is needed. There are many other *MVA* methods which are combined with *MIR* spectroscopy and used by other researchers. There is a series of previous studies which uses Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) as *MVA* methods and, when these methods are combined with *IR* spectroscopy, are reported as providing excellent accuracy with great potential for identification and classification of food items. A study by Strong et al. (2017) used both unsupervised PCA and supervised LDA as tools in their multivariate analysis (*MVA*) and their work successfully investigates a number of biological phenomena obtained from their *MIR* studies. In addition, Llabjani et al. (2011) assert that the best way to reduce IR spectra dimensionality is to use *MVA* (PCA and LDA). Because IR spectra contain a large quantity of information, they can be considered to inhabit a large multidimensional space and thus a dimensionality reduction method is needed.

In this study, we propose a method of spectroscopic analysis that allows subjects to discriminate between two food sample spectra using audio. We introduce a spectral pre-processing method, followed by feature extraction, and dimensionality reduction, and finally classify the spectra by synthesising sound from the spectra. Analysis time can be reduced, and real-time feedback is possible using feature extraction. Such *sonification* becomes a supporting tool which participants can use to facilitate the analysis of large spectroscopy datasets and can readily detect any abnormality in the spectra. The aim of our model is to provide data feedback via a sound signal which then frees a user to concentrate on other aspects of the task. Furthermore, the tool is then also able to reduce the burden of subjective analysis from users.

## 2 Materials and methods

### 2.1 Dataset

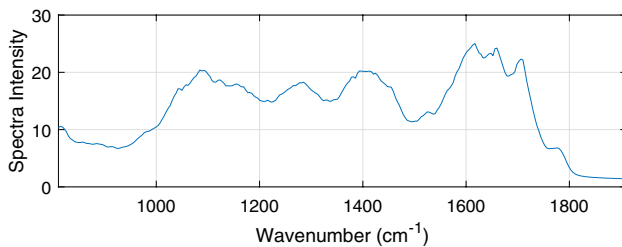
Two real-life datasets were analysed using our model, to illustrate the performance of our developed sonification

algorithm when applied to spectroscopy data. These were a *coffee dataset* and a *fruit puree dataset*, both datasets are open-source, available via the internet, are collected by an Fourier Transform Infrared (*FTIR*) spectrometer and were provided by Analytical Sciences Unit of the Institute of Food Research, UK Quadram Institute Bioscience (2007). Henríquez and Ruz chose six spectral datasets for their studies, in which they analysed the use of near-infrared spectroscopy Henríquez and Ruz (2019). They reduced the noise using Extreme Learning Machines (*ELM*). These six datasets were used to evaluate the performance of their developed models. Among those six datasets, two of the datasets, both the *coffee dataset* and the *fruit puree dataset*, are used in our studies. In addition, Zhongzhi also selected four spectral datasets to evaluate their novel development of a kernel support vector machine (*k-SVM*) to distinguish different foods Zhongzhi (2019). Similarly, among these four spectroscopic datasets, two of the datasets are used in our work.

#### 2.1.1 Coffee dataset

The first food analysis dataset contains *FTIR* spectra of food items, collected by Briandet et al. Briandet et al. (1996). The spectra were collected via attenuated total reflection (*ATR*) and diffuse reflectance infrared Fourier transform (*DRIFT*) methods. This study is focused on the classification of two different types of instant freeze-dried coffee, *Arabica* and *Robusta*. With regard to its chemical composition, coffee is one of the most complex food commodities. *Arabica* coffee accounts for 90% the world coffee production while *Robusta* coffee accounts for approximately 9% of world coffee production Briandet et al. (1996). *Arabica* beans are generally regarded as having a superior flavour which is more pronounced than that of the *Robusta* Coffee bean. *Arabica* coffee thus has a higher price and is more highly valued. There is hence a tendency to adulterate *Arabica* coffee with *Robusta* coffee because of the price difference. Therefore, it is important to identify the variety of beans in the coffee by using a classification method in order to avoid accidental or fraudulent mislabelling. The original paper by Briandet et al. Briandet et al. (1996) explores whether the datasets can be quantified as a mixture of the two, or whether the datasets can be authenticated alone. Figure 1 shows a typical *FTIR* spectrum Briandet et al. (1996). The spectrum has been area-normalised and baseline corrected. In the range  $1800\text{ cm}^{-1}$  to  $1900\text{ cm}^{-1}$ , Fig. 1 shows that detector noise is almost negligible.

The raw dataset consists of 56 samples in total which consists of 29 *Arabica* samples and 27 *Robusta* samples. Datasets with a nominal resolution of  $3.85\text{ cm}^{-1}$  were collected in the  $810.548\text{ cm}^{-1}$  to  $1910.644\text{ cm}^{-1}$  region of the spectra, and are shown in Fig. 2. The spectra of both species appear quite similar in shape and detail even though



**Fig. 1** A typical *FTIR* spectrum from sample numbered 2, *Arabica*, Class A of the *coffee dataset* Briandet et al. (1996)

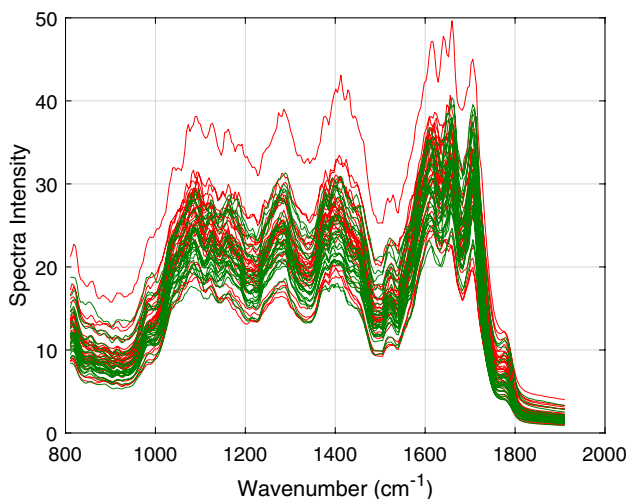
both species exhibit physical and chemical composition differences. There are some small differences which can be identified on visual inspection, particularly in the ranges of  $1000\text{ cm}^{-1}$  to  $1300\text{ cm}^{-1}$  and  $1600\text{ cm}^{-1}$  to  $1800\text{ cm}^{-1}$ . It is however very difficult to distinguish between the two spectra by simply looking at the spectra i.e. by visualisation alone. Therefore, because of the complexity of the spectra, multivariate statistical techniques is used to discriminate between the two species.

### 2.1.2 Fruit puree dataset

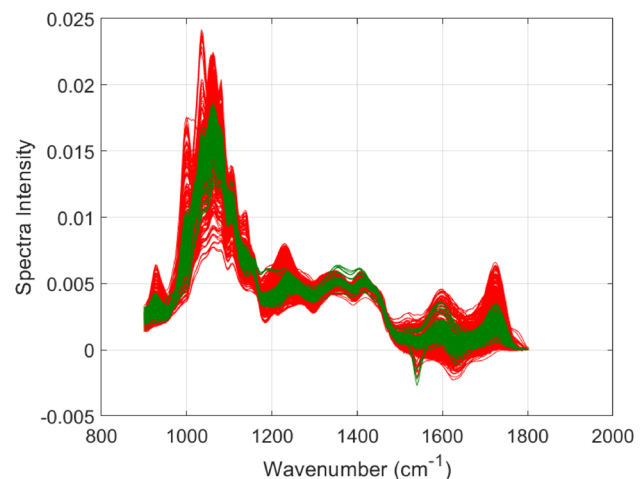
A larger and more complex dataset (*fruit puree dataset*) is used for the purpose of validating the feasibility of the sonification algorithm developed using the *coffee dataset*, by applying them to different datasets. The second dataset contains a collection of 983 mid-infrared spectra in total, split between one of two classes: pure strawberry data

(authentic samples) and non-pure strawberry, which latter contains adulterated strawberries and other fruits Holland et al. (1998). There are 632 samples of pure strawberry data (authentic samples) which we refer to as Class A and 351 samples of non-pure strawberry data (adulterated strawberries and other fruits) which we refer to as Class B. A nominal resolution of  $8\text{ cm}^{-1}$  was used, and a total of 256 interferograms were co-added and a triangular apodisation used prior to Fourier transformation Holland et al. (1998), Defernez and Wilson (1995). The data were collected using *FTIR* spectroscopy with *ATR* sampling. The raw data matrix size is  $983 \times 235$  (983 samples and 235 wavenumbers) and has a nominal resolution of  $3.86\text{ cm}^{-1}$ , collected in the  $899.327\text{ cm}^{-1}$  to  $1802.564\text{ cm}^{-1}$  range. By looking at the plot of the *fruit puree dataset* in Fig. 3, it is difficult to accurately determine visually whether there is any systematic difference between the pure strawberry and the non-pure strawberry data samples.

From the *fruit puree dataset*, pure strawberry spectra (Class A) sample numbers are from A65 to A89, A98 to A135, A182 to A227, A248 to A257, A400 to A424, A433 to A469, A515 to A560, A579 to A587, A726 to A749, A757 to A793, A838 to A882 and A899 to A907 whilst non-pure strawberry spectra (Class B) sample numbers are from B1 to B64, B90 to B97, B136 to B181, B228 to B247, B258 to B399, B425 to B432, B470 to B514, B561 to B578, 588 to B725, B750 to B756, B794 to B837, B883 to B898 and B908 to B983. Due to the fact that the original dataset consists of 983 samples of data (which is considered a large dataset), a smaller subset of the data was, somewhat



**Fig. 2** 56 full mid-infrared (*MIR*) spectra of authenticated freeze-dried coffee samples: Red colour plots represent *Arabica* species (Class A), and green colour plots represent *Robusta* species (Class B), with 29 and 27 of each respectively. The vertical axis displays the intensity (offset), and the horizontal axis is the wavenumber,  $\text{cm}^{-1}$  Briandet et al. (1996)



**Fig. 3** 983 full mid-infrared (*MIR*) spectra of fresh fruit puree samples: green colour plots represent 632 samples of pure strawberry data (authentic samples) which we refer to as Class A, and red colour plots represent 351 samples of non-pure strawberry data (adulterated strawberries and other fruits) which we refer to as Class B. The vertical axis displays the intensity (offset), and the horizontal axis is the wavenumber ( $\text{cm}^{-1}$ ) Holland et al. (1998)

arbitrarily, selected for analysis and display and to consist of a small number of spectra with a high relative intensity, a small number of spectra with a moderate (or medium) relative intensity and finally a small number of spectra of low relative intensity. These data should demonstrate typical behaviour patterns without the time constraints of using the entire large dataset. This can also avoid overloading the data during analysis.

### 2.1.3 Model

This spectroscopic analysis model is written in the programming language, *MATLAB* (The Math Works Inc, Natick, Massachusetts, USA), and the statistical tool provided in release R2017b has been used because of its powerful data processing capabilities and scalabilities. The data processing was implemented on a personal computer.

## 3 Model

Vibrational spectroscopy is simple to implement because it requires minimum sample preparation. However, it is not suitable for real-time spectral analysis as the process is time consuming. Fig. 4 shows a brief overview of our proposed method. The processing flows are described below.

### 3.1 Spectral pre-processing

Spectral pre-processing is a process whereby all data channels are pre-analysed for numerical range, linear and exponential averages, and user-customisable analysis for the purpose of convenience de Campo et al. (2004). These are essential steps which need to be performed prior to spectral analysis to ensure there is a commonality between the spectra. The aim is to improve the characteristics sought in the spectra and to improve also the data quality. It also

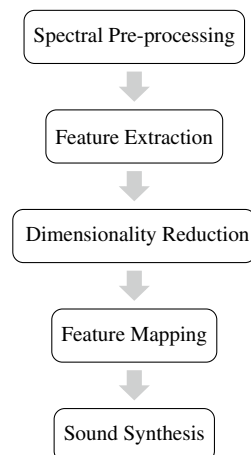
aims to reduce the noise in the data and remove the physical phenomena (noise) from the spectra in order to improve the classification between 2 samples of data. This, can be achieved by choosing suitable pre-processing techniques. However, according to Rinnan et al., applying too many pre-processing methods or selecting a wrong method will lead to a bad outcome and which will delete valuable information Rinnan et al. (2009). Another difficulty is that many pre-processing techniques are now being developed and are freely available to choose from, and a combination of different pre-processing techniques can also be used. A variety of filters and techniques are used in pre-processing and noise reduction. One of the non-linear approaches is called *morphological filter* and is based on a combination of median filtering and classical gray-scale morphological operators. Such approach has been applied successfully to biomedical signal and image processing Sedaaghi et al. (2001). Another approach described by Stone Stone (2006, 2004) is called *Blind Source Separation* and performs blind separation of the mixtures based on temporal predictability. This approach has received much attention because of its low complexity, the used of second-order statistics techniques and its efficiency Khosravy et al. (2020). Thus, a robust methodology is needed to solve this problem. Spectral pre-processing is an important step, which has a significantly positive effect on the process of discrimination between two classes and the quality of spectral feature extraction.

#### 3.1.1 Savitzky–Golay (S–G) smoothing filtering

This smoothing method is widely used in spectrum pre-processing to eliminate the noise presented in a spectrum, especially high frequency noise. We use a smoothing filter, called a Savitzky–Golay filter, which is a popular tool used to clean up a spectrum without introducing significant distortion. One of the advantages of Savitzky–Golay filter is that it does not filter out high frequency content along with the noise as compared to a standard averaging finite impulse response (FIR) filter. The Savitzky–Golay filter is also known as a least squares smoothing filter, or digital smoothing polynomial filter, as it minimises the least-squares error in fitting to the successive subset of adjacent data points. It reduces the noise while maintaining the height and shape of waveform peaks Savitzky and Golay (1964).

For the purposes of comparison, Fig. 5a–c shows spectra both with and without pre-processing, obtained using the sample numbered 2 of the *Arabica* (Class A of the *coffee dataset*). A similar setting of Savitzky–Golay smoothing filtering is applied to the *fruit puree dataset* as shown in Fig. 6a–b. As shown, the spectrum is smoother and also there are some changes in the spectrum due to different pre-processing techniques used. Thus, the performance of

**Fig. 4** The flow chart of explaining our sonification algorithm



sonification algorithm in classification is strongly influenced by the pre-processing method used.

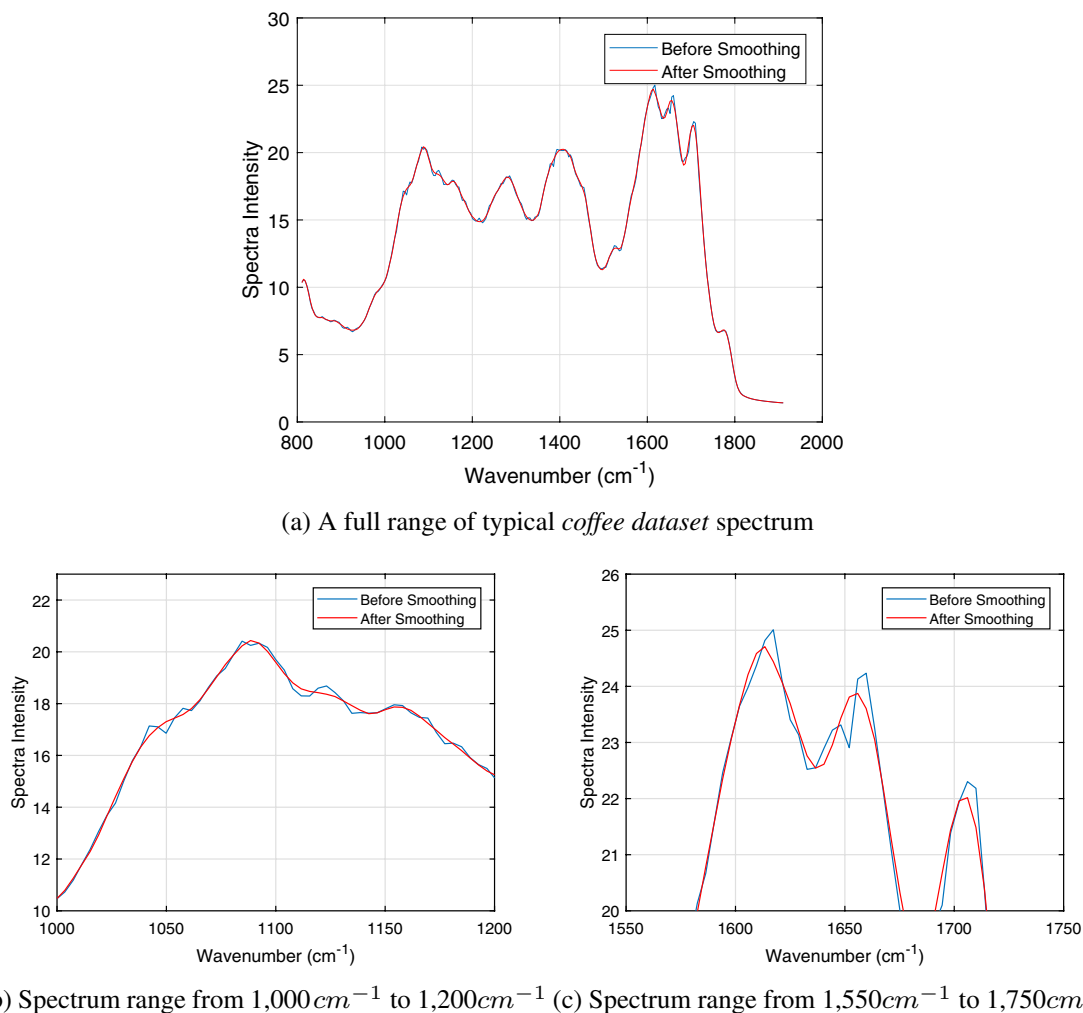
### 3.1.2 Segmentation

A segmentation method is used in spectral data analysis to divide a spectrum into several pieces, known as *segments*. Each segment contains features that can be readily analysed using such a method. A particular segment can be focused on, if that segment contains important features of spectral data. In this work, an entire spectrum was divided into four segments based on the prominence of peaks detected. Particular features from each segment were extracted and the procedure will be described and presented in the following section. Figure 7 shows an example of the segmentation method applied to a spectrum.

### 3.2 Feature extraction

Spectral features provide contextual information, and so contain much useful information. Feature extraction is a necessary processing step to achieve a discriminant analysis of spectroscopic data, and further improve the efficiency of *FTIR* spectral data. A study of feature extraction and analysis of Raman spectroscopic analysis of dairy products based on Raman peak intensity is described by Zhang et al. and these authors demonstrated clearly that spectral feature extraction can improve the efficiency of Raman spectral data as a method of classification Zhang et al. (2019).

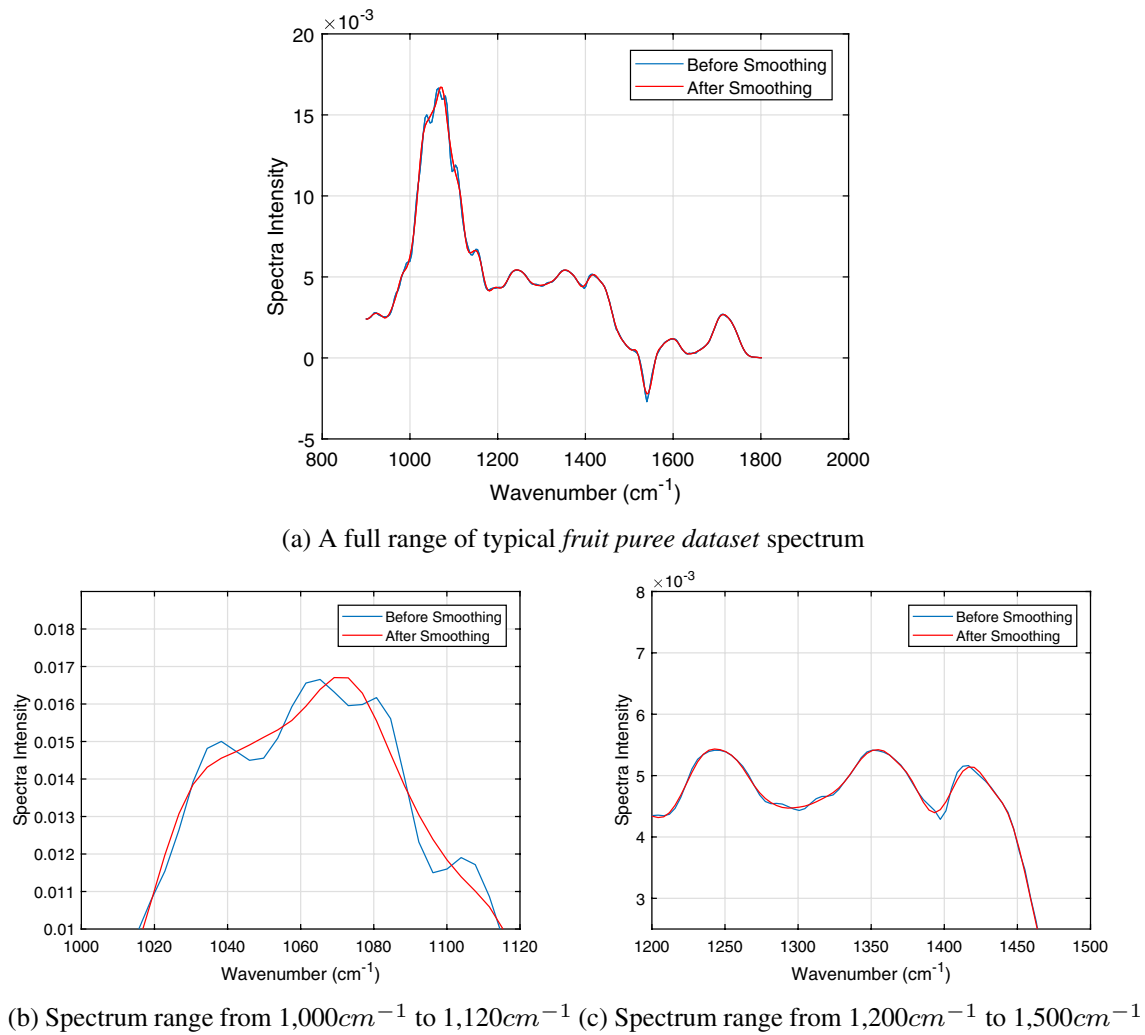
The most salient properties of the absorption spectra are extracted for the purpose of maximising the classification accuracy. Firstly, we must seek the peaks or the local maxima of the spectrum. However, there are multiple peaks within any one spectrum and the peak that



**Fig. 5** Spectra before and after Savitzky-Golay smoothing filtering using sample numbered 2 of the *Arabica* (Class A of the *coffee* dataset) Briand et al. (1996). **a** A full range of typical *coffee* data-

set spectrum. **b** Spectrum range from  $1000$  to  $1200\text{ cm}^{-1}$ . **c** Spectrum range from  $1550$  to  $1750\text{ cm}^{-1}$





**Fig. 6** Spectra before and after Savitzky–Golay smoothing filtering using sample numbered 65 of the pure strawberry samples (Class A of the *fruit puree dataset*) Holland et al. (1998). **a** A full range of

typical *fruit puree dataset* spectrum. **b** Spectrum range from 1000 to 1120  $\text{cm}^{-1}$ . **c** Spectrum range from 1200 to 1500  $\text{cm}^{-1}$

stands out is extracted by measuring its intrinsic height/amplitude and its location relative to other peaks. Because some peaks are very close to each other, we have to determine the prominence of any given peak. A peak is determined by the discrete sampling of data at intervals, ( $\delta$ ), of wavenumber. Firstly, all the prominent peaks, or the local maxima of a spectrum were detected. In the process of prominent peak detection, we used a conditional statement in which the value of minimum peak prominence is increased until the number of prominent peaks detected reaches the value of 4. The whole process of prominent peak detection using a spectrum obtained from the *coffee dataset* (Sample numbered 2, Arabica, Class A) is shown in Fig. 8. A total of 28 peaks was detected in sample numbered 2 of the *Arabica* (Class A of the *coffee dataset*) as shown in Fig. 8a. The prominent peaks are sorted from the highest to the lowest. Thus, the minimum peak prominence

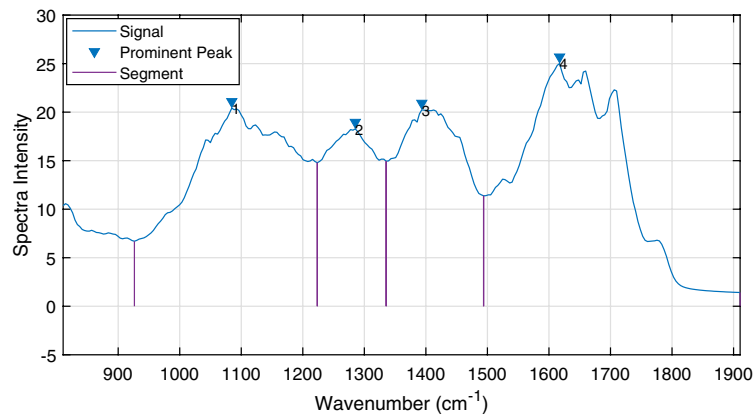
needs to be increased until the number of prominent peaks is 4 as shown in Fig. 8c using a conditional statement.

Each spectrum is then divided into 4 segments based on the 4 prominent peaks detected in the earlier process. Important features are extracted from each segment in sequence. The sonification algorithm is then followed by extraction of five features in each segment which are the peak amplitude, the peak frequency, the peak width, the positive gradient and the negative gradient of a chord joining the prominent peak to the two adjacent troughs as shown in Fig. 9.

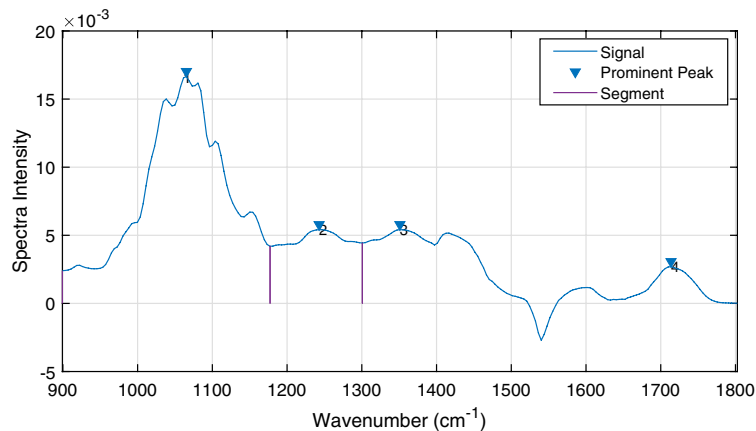
### 3.2.1 Peak amplitude

The peak amplitude refers to the intensity (relative) value of a spectrum at which the amplitude has a peak. The values were measured using the prominent peak detection method we have discussed.

**Fig. 7** Spectral pre-processing method: Segmentation. **a** Segmentation of a spectrum from Sample numbered 2 of the *Arabica* (Class A of the *coffee dataset*) Briandet et al. (1996). **b** Segmentation of a spectrum from Sample numbered 65 of the pure strawberry samples (Class A of the *fruit puree dataset*) Holland et al. (1998)



(a) Segmentation of a spectrum from Sample numbered 2 of the *Arabica* (Class A of the *coffee dataset*) [20].



(b) Segmentation of a spectrum from Sample numbered 65 of the pure strawberry samples (Class A of the *fruit puree dataset*) [21].

### 3.2.2 Peak frequency

The peak frequency refers to wavenumber value of a spectrum at which the amplitude has a peak.

### 3.2.3 Peak width

The peak width is returned as a vector of real numbers, which is the length of the length of a horizontal reference line at one half of the peak height. The border between peaks was defined by the horizontal position of the lowest valley between two peaks.

### 3.2.4 Upward slope/positive gradient

We use the Cartesian coordinate  $x$  to define the position of a point on the spectrum whose height is  $y$ . Associated with a prominence peak (or local maximum) at  $(x_2, y_2)$  is a point  $(x_1, y_1)$ , which marks the end of the segment lying to the left of the peak, and a point  $(x_3, y_3)$  which marks the end of the

segment lying to the right of this peak. If the points  $(x_1, y_1)$  and  $(x_2, y_2)$  are joined by a straight line, it will have a (positive) gradient  $m_{pg}$  given by

$$m_{pg} = \frac{\Delta y}{\Delta x} = \frac{(y_2 - y_1)}{(x_2 - x_1)}. \quad (1)$$

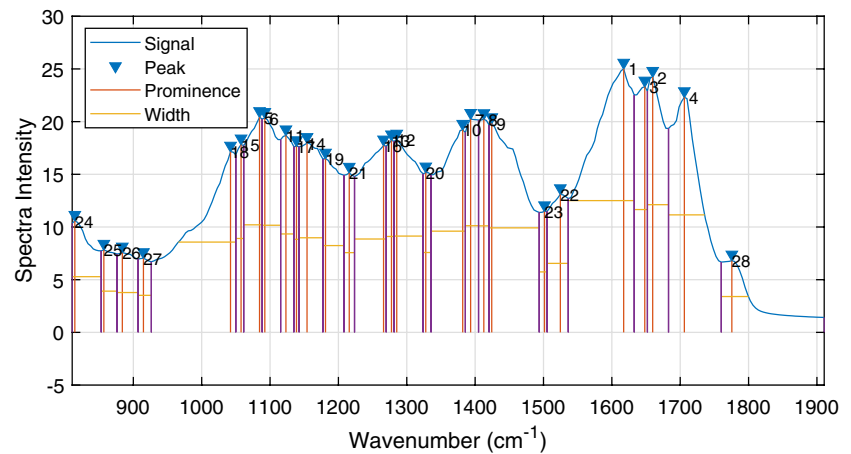
### 3.2.5 Downward slope/negative gradient

Similarly the straight line passing through  $(x_2, y_2)$  and  $(x_3, y_3)$  will have a (negative) gradient  $m_{ng}$  given by

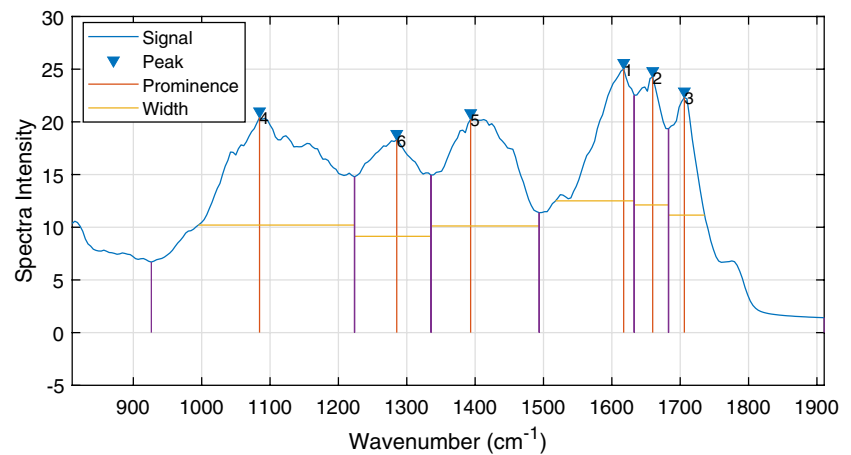
$$m_{ng} = \frac{\Delta y}{\Delta x} = \frac{(y_3 - y_2)}{(x_3 - x_2)}. \quad (2)$$

These two numbers  $m_{pg}$  and  $m_{ng}$  constitute two of the features we associate with each segment and, alongside the peak height, peak position and peak width, constitute the five features we associate with a given segment. Since each spectrum is segmented into four segments, and each segment has five associated features, we associate a total of twenty extracted features with each entire spectrum. Figure 9 shows

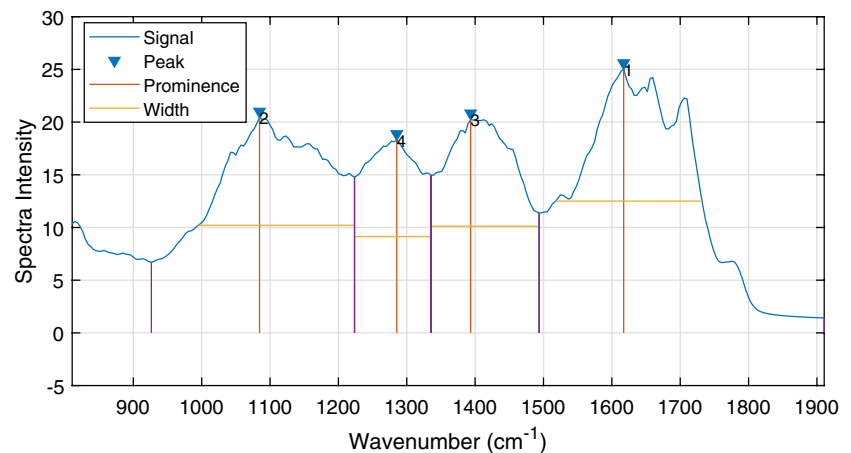
**Fig. 8** The process of peak analysis on a spectrum from sample numbered 2 of the *Arabica* (Class A of the *coffee dataset*) Briandet et al. (1996). **a** Minimum peak prominence value is 0 and the number of prominent peaks detected is 28. **b** Minimum peak prominence value is 1 and the number of prominent peaks detected is 6. **c** Minimum peak prominence value is 3 and the number of prominent peaks detected is 4



(a) Minimum peak prominence value is 0 and the number of prominent peaks detected is 28



(b) Minimum peak prominence value is 1 and the number of prominent peaks detected is 6



(c) Minimum peak prominence value is 3 and the number of prominent peaks detected is 4

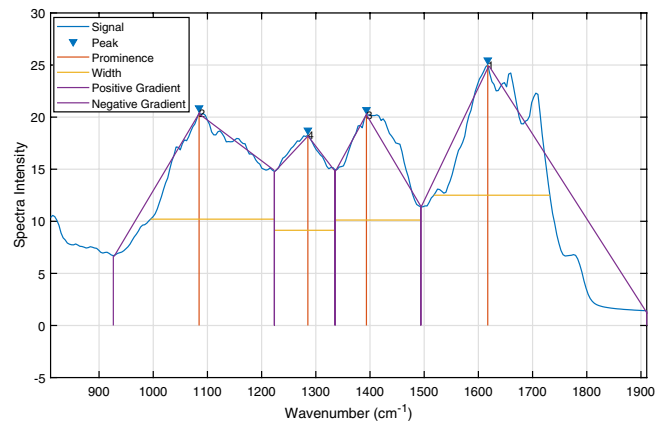
all the features being detected in the spectrum which include the peak amplitude, peak frequency, peak width, positive gradient and negative gradient in one of the spectra from the *coffee dataset* and the *fruit puree dataset*.

### 3.3 Dimensionality reduction

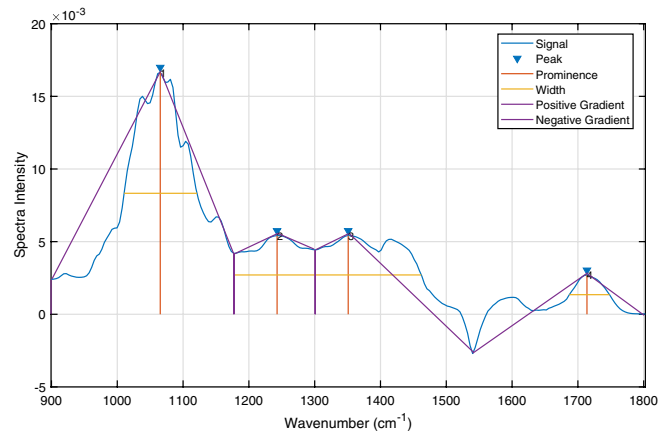
To reduce the number of features, we compare the Principal Component Analysis (PCA) and the Linear



**Fig. 9** Extraction of the features. **a** Features extraction of a spectrum from sample numbered 2 of the *Arabica* (Class A of the *coffee dataset*) Briand et al. (1996). **b** Features extraction of a spectrum from sample numbered 65 of the pure strawberry samples (Class A of the *fruit puree dataset*) Holland et al. (1998)



(a) Features extraction of a spectrum from sample numbered 2 of the *Arabica* (Class A of the *coffee dataset*) [20].



(b) Features extraction of a spectrum from sample numbered 65 of the pure strawberry samples (Class A of the *fruit puree dataset*) [21]

Discriminant Analysis (LDA) methods. PCA is the most popular and classical method to the unsupervised approach of dimensionality reduction. LDA is particularly suitable for a two-class classification task and matches the accuracy of our work in discriminating two classes of datasets. Both algorithms require relative low computing power to operate when compared to other methods.

In our experiment, we associate a total of twenty features with each spectrum. Using the data from these twenty features, the algorithm then applies a dimensionality reduction method using *PCA* in order to reduce the data to a final four outputs. After finding the principal components, these four outputs will be used as input parameters for the synthesis process. This process is called a ‘many-to-one’ mapping in parameter mapping sonification Grond and Berger (2011). Dimensionality reduction plays a significant role in parameter mapping techniques by continuously controlling the transformation of dataset into sound.

### 3.3.1 Principal component analysis (PCA)

The origin of principal component analysis (PCA) can be traced back to Pearson (1901) who used it for the problem of data fitting. *PCA* is one of the most common *MVA* methods used to reduce the dimensionality of a dataset. *MIR* spectra data are considered as high dimensional data, and with the use of *PCA*, such data can be projected onto a lower-dimensional space. By finding combinations of variables, *PCA* is able to extract relevant and useful information from multivariate data. Reducing the dimensionality of the feature space takes no account of the class labels.

### 3.3.2 Linear discriminant analysis (LDA)

Another widely used discriminant method in chemometric applications is linear discriminant analysis (LDA), first proposed by Fisher (1936). LDA chooses a direction in the space of the data and projects the data onto a line which

is chosen to maximise the distance between the projected means of the two classes (Class A and Class B), and at the same time minimise the sum of the projected spreads of each class of data about the respective projected mean of that class of data, thereby guaranteeing maximum class separation.

### 3.4 Feature mapping

To link the output of the dimensionality reduction process with the parameters of sound synthesis, the features need to be scaled into appropriate parameter ranges, which lie in the range of 1 to 100. The sound synthesis parameters are the carrier amplitude, modulator amplitude, carrier frequency and modulator frequency, which latter is chosen to lie within the range of human hearing when generated as audio output.

### 3.5 Sound synthesis

To synthesize the input samples, Amplitude Modulation (AM) and Frequency Modulation (FM) synthesis are chosen. Both synthesis methods are relatively easy to use and design, and yet still provide a variety of sonic outcomes. Both algorithms require relative low computing power to operate as compared to other methods. These methods are able to generate a rich variety of sounds with a small number of controllable parameters.

#### 3.5.1 Amplitude modulation (AM)

In AM synthesis, a carrier signal is fed into a modulator, which is a second oscillator whose function is to modulate the amplitude (or volume) of the first oscillator. Both the carrier and the modulator are oscillators. However, the carrier can be another form of signal such as a vocal or instrumental input. If the AM synthesis uses a low frequency below the threshold of human hearing (approximately 20Hz), then the audio signal associated with the carrier will then appear to become alternately louder and then quieter. The amplitude modulated signal can be expressed as:

$$S_{AM}(t) = [A_c + A_m \sin(2\pi f_m t)] \sin(2\pi f_c t). \quad (3)$$

In our study, we control four parameters of the AM synthesis which are the carrier frequency  $f_c$ , carrier amplitude  $A_c$ , modulator frequency  $f_m$  and modulator amplitude  $A_m$ . The four outputs from the PCA in our sonification algorithm can then be mapped into the these four parameters of the AM.

#### 3.5.2 Frequency modulation (FM)

FM synthesis is the periodic variation of frequency of a signal. FM synthesis uses one wave to rapidly modulate the

frequency of another wave thus generating a collection of entirely new frequencies. The principle is based on using at least two oscillators, one of which is a carrier and the other is a modulator. Modulating the frequency of one with the frequency of the other, generates a complex waveform which is rich in harmonics. The frequency modulated signal can be expressed as:

$$S_{FM}(t) = A \sin[2\pi f_c t + \beta \sin(2\pi f_m t)]. \quad (4)$$

In our study, we control four parameters of the FM synthesis: the carrier frequency  $f_c$ , carrier amplitude  $A$ , modulator frequency  $f_m$  and modulator amplitude  $\beta$ . In a similar manner to the mapping of the AM synthesis, the four outputs from the PCA in our sonification algorithm can be mapped onto these four parameters. The audio clips can then be used to distinguish between the two classes because it is able to create clear differences of timbre.

## 4 Experimental method and results

### 4.1 Experimental setup

To evaluate the proposed spectroscopic analysis model's ability to produce audible differences between the two classes, a listening test is setup using The Web Audio Evaluation Tool Jillings et al. (2016) which is publicly available online. This tool was used to build and run listening tests in the browser. To subjectively evaluate the synthesized audio samples, an ABX test Clark (1982) is used as described below.

This latter test is a method for comparing two choices of audio samples with a reference audio sample in order to identify the difference between the two audio samples. In the listening test, each participant is given two known audio samples, referred to as stimuli, one from Class A (*Arabica*) and one from Class B (*Robusta*) respectively, and these are then followed by one unknown audio sample, reference (X), that is randomly selected from either Class A or Class B. Two stimuli (Class A and Class B of the dataset) are thus presented along with a reference (X) stimulus (either from Class A or Class B) and the participant has to select which of the stimuli (either Class A or Class B) most closely resembles the reference (X) stimulus. For example, a typical listening test for Stimulus 1 is labelled as A5, B42 and B33 as shown in Table 1, where A5 represents the audio sample numbered 5 from Class A whereas B42 represents the audio sample numbered 42 from Class B. Finally B33 is the reference audio sample numbered 33 of Class B. When the listening test is setup, it has ten stimuli. The participant is then required to identify (X) as either A or B.

**Table 1** Experimental setup for *Experiment 1* using the *coffee dataset*

Stimulus number	Class A sample numbers	Class B sample numbers	Reference (X) Sample numbers
Stimulus 1	A5	B42	B33
Stimulus 2	A10	B33	B44
Stimulus 3	A17	B30	A7
Stimulus 4	A9	B30	B41
Stimulus 5	A27	B50	A28
Stimulus 6	A26	B47	A27
Stimulus 7	A2	B47	A28
Stimulus 8	A18	B34	B33
Stimulus 9	A22	B38	A8
Stimulus 10	A5	B35	B52

The test is designed so that the reference sample (X), cannot be the same sample number as the Class A sample number and that Class B sample number. Thus, for stimulus 1 as shown in Table 1, the Class A sample is A5, the Class B sample is B42 and the reference (X) sample is not permitted to be either of these, but *is* chosen from the same category of intensity. As shown in Table 1, stimulus 1, category reference (X) is chosen to be B33 etc. The reference (X) audio sample chosen for each stimulus is split 50% from Class A samples and 50% from Class B samples for greater validity of the test.

Table 1 shows the experimental setup using *coffee dataset* and Table 2 shows the experimental setup using *fruit puree dataset*

Figure 10 below shows interface front page of the listening test, and Fig. 11 shows the interface of the ABX test.

The data presented for the listening test in for this particular study is available at the sites listed below:

- (i) <http://dmtlab.bcu.ac.uk/hsein/waet/test.html?url=tests/sonificationlisteningtest2.xml>  
for *Experiment 1 (Part 1)*
- (ii) <http://dmtlab.bcu.ac.uk/hsein/waet/test.html?url=tests/sonificationlisteningtest3.xml>  
for *Experiment 1 (Part 2)*
- (iii) <http://dmtlab.bcu.ac.uk/hsein/waet/test.html?url=tests/sonificationlisteningtest5.xml>  
for *Experiment 2 (Part 1)*
- (iv) <http://dmtlab.bcu.ac.uk/hsein/waet/test.html?url=tests/sonificationlisteningtest6.xml>  
for *Experiment 2 (Part 2)*

## 4.2 Participants

A total of 21 English speaking participants with normal hearing, aged between 18 and 40, participated in the

**Table 2** Experimental setup for *Experiment 2* using the *fruit puree dataset*

Stimulus number	Class A sample numbers	Class B sample numbers	Reference (X) sample numbers
Stimulus 11	A65	B2	A70
Stimulus 12	A79	B10	B13
Stimulus 13	A 128	B13	B27
Stimulus 14	A 518	B2	A529
Stimulus 15	A 204	B36	B49
Stimulus 16	A 524	B64	A529
Stimulus 17	A 841	B49	A70
Stimulus 18	A 857	B56	A743
Stimulus 19	A 204	B49	B50
Stimulus 20	A 547	B58	B60

listening test. All participants have self-reported normal hearing and are fluent in English. The participants are non-specialists in spectroscopic analysis, and therefore were unable to discriminate between the two classes by looking at the spectrum. The listening test consisted of a total of 20 stimuli of test sounds. Each participant listened to 20 different stimulus test sounds and was asked to decide which sound most closely resembled the reference sound.

## 4.3 Experimental results and analysis

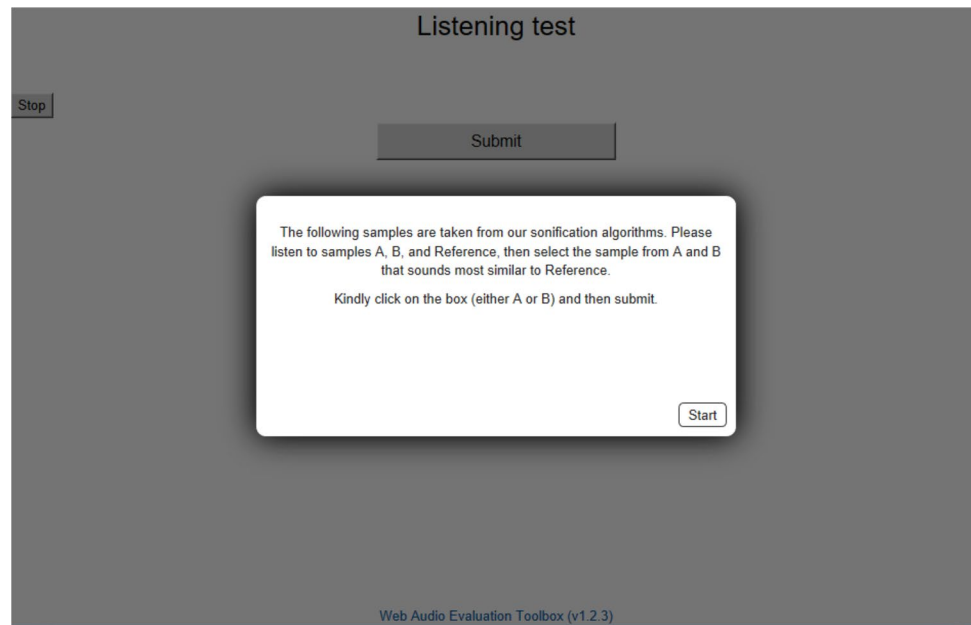
Two experiments were conducted in this study using the two sound synthesis methods. *Experiment 1* uses the *coffee dataset* for the initial implementation of the sonification algorithm. *Experiment 2* uses the *fruit puree dataset*, a second, larger and more complex dataset, and is used to further validate the results and serve as supportive evidence.

### 4.3.1 Experiment 1 (Part 1): Coffee Dataset

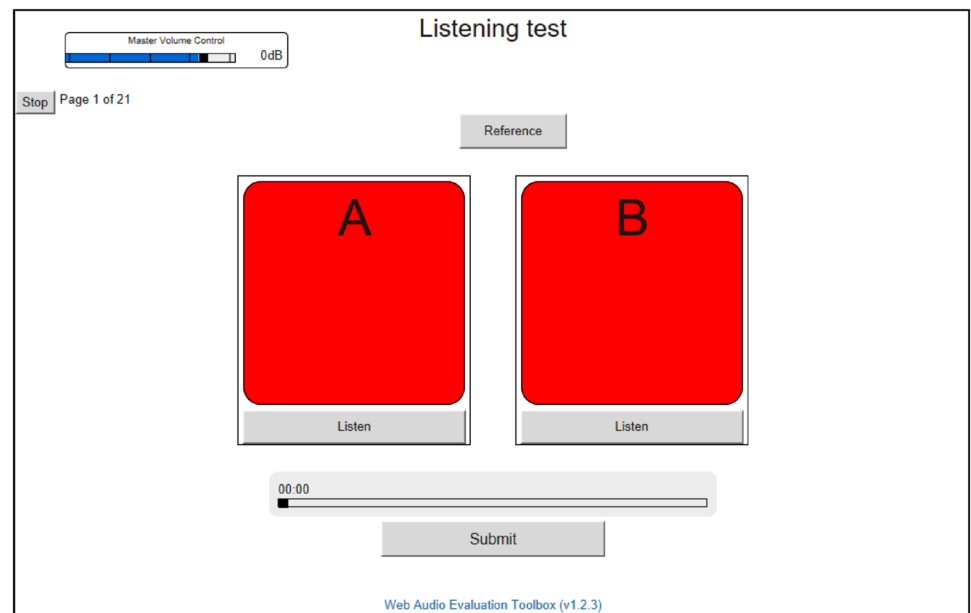
The results of the classification accuracy experiment for both AM and FM methods are presented in Table 3. The AM method is tested with ten stimuli and the FM method is tested with another ten stimuli. Thus, twenty stimuli in total are used. Both audio stimuli are generated from the same class and same sample number, the only difference is the method, which is either AM or FM.

The results show that the mean accuracy obtained using the AM method of synthesis is 50.48% whereas the mean accuracy obtained using the FM method of synthesis is 93.81%. The median for AM is 50.00% and the median for FM is 95.24% as shown in Fig. 12. This strongly suggests that classification accuracy for FM is again significantly higher than for AM using the *coffee dataset*.

**Fig. 10** The first interface of the listening test



**Fig. 11** Listening test interface



A non-parametric Wilcoxon Signed Rank Test was conducted, and which indicated a statistically significant difference in mean classification accuracy between the AM sound synthesis ( $Mdn = 50.00\%$ ) and FM sound synthesis ( $Mdn = 95.24\%$ ), with  $Z = 2.803$ ,  $p = 0.005$ , and  $r = 0.63$ .

#### 4.3.2 Experiment 1 (Part 2): Coffee Dataset

The recommended method identified in *Experiment 1 (Part 1)* is used, together with the addition of a dimensionality

reduction method called LDA and a comparisons is made with PCA. The second test which uses the *PCA* method is not repeated again however, but is taken from *Experiment 1 (Part 1)*. The results of the classification accuracy experiment for both the PCA and the LDA methods are presented in Table 4. Stimulus 1 is processed with the PCA method and is also processed with the LDA method.

Overall, the mean accuracy obtained using PCA is 93.81% whereas the mean accuracy obtained using the LDA is 37.50%. The classification accuracy using the PCA method is higher than the accuracy obtained using the LDA

**Table 3** Classification accuracy of *Experiment 1 (Part 1)* using the *coffee dataset*

Stimulus number <sup>a</sup>	Classification accuracy (%)	
	AM	FM
Stimulus 1 (A5, B42; B33)	33.33	100.00
Stimulus 2 (A10, B33; B44)	57.14	80.95
Stimulus 3 (A17, B30; A7)	57.14	85.71
Stimulus 4 (A9, B30; B41)	42.86	90.48
Stimulus 5 (A27, B50; A28)	42.86	95.24
Stimulus 6 (A26, B47; A27)	57.14	100.00
Stimulus 7 (A2, B47; A28)	42.86	100.00
Stimulus 8 (A18, B34; B33)	61.90	95.24
Stimulus 9 (A22, B38; A8)	38.10	100.00
Stimulus 10 (A5, B35; B52)	71.43	90.48
Mean (%)	50.48	93.81
Standard deviation (%)	12.13	6.75
Median (%)	50.00	95.24

<sup>a</sup>Stimulus number (Class A sample numbers, Class B sample numbers; Reference (X) sample numbers)

method which we have used in this study of dimensionality reduction.

The LDA method fails to find the projection defining the LDA space when the discriminatory information does not lie in the means of classes, because the classes are non-linearly separated. This is known as the *singularity problem* in spectroscopic data which Krzanowski et al. (1995) said that the

LDA requires the total scatter matrix to be non-singular. As a result, for the LDA method, the audio signal generated from the FM synthesis does not have much differences in timbre, and this leads to low classification accuracy. On the other hand, PCA finds the principal components and these four outputs will be used as input parameters for the synthesis process, and create clear differences of timbre.

A non-parametric Mann–Whitney *U* Test with an alpha of 5% was conducted because the results obtained in *Experiment 1* are non-parametric and not normally distributed, and this test indicated a statistically significant difference in mean classification accuracy between PCA ( $Md = 95.24\%$ ) and LDA ( $Md = 37.50\%$ ), with  $U = 0.000$ ,  $Z = -3.801$ ,  $p < 0.001$ ,  $r = 0.85$ .

#### 4.3.3 Experiment 2 (Part 1): *Fruit Puree Dataset*

The results of the classification accuracy experiment for both AM and FM methods are presented in Table 5. The results show that the mean of the classification accuracy for AM is 57.62% and the mean of the classification accuracy for FM is 88.57%. The median for AM is 59.52% and the median for FM is 90.48% as shown in Fig. 12. This strongly suggest that the classification accuracy for FM is significantly higher than for AM using the *fruit puree dataset*.

An independent-samples *t*-test was conducted, which indicated a significant difference between the mean values using PCA ( $M = 88.57\%$ ,  $SD = 11.49\%$ ) and LDA ( $M = 42.50\%$ ,  $SD = 10.36\%$ ); with  $t(18) = 9.418$ ,  $p < 0.001$ . The

**Table 4** Classification accuracy of *Experiment 1 (Part 2)* using the *coffee dataset*

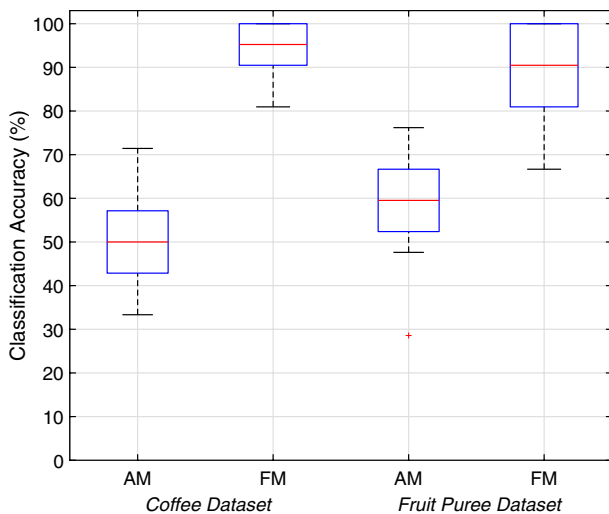
Stimulus number <sup>a</sup>	Classification accuracy (%)	
	PCA	LDA
Stimulus 1 (A5, B42; B33)	100.00	41.67
Stimulus 2 (A10, B33; B44)	80.95	37.50
Stimulus 3 (A17, B30; A7)	85.71	45.83
Stimulus 4 (A9, B30; B41)	90.48	50.00
Stimulus 5 (A27, B50; A28)	95.24	25.00
Stimulus 6 (A26, B47; A27)	100.00	29.17
Stimulus 7 (A2, B47; A28)	100.00	41.67
Stimulus 8 (A18, B34; B33)	95.24	33.33
Stimulus 9 (A22, B38; A8)	100.00	37.50
Stimulus 10 (A5, B35; B52)	90.48	33.33
Mean (%)	93.81	37.50
Standard Deviation (%)	6.75	7.61
Median (%)	95.24	37.50

<sup>a</sup>Stimulus number (Class A sample numbers, Class B sample numbers; Reference (X) sample numbers)

**Table 5** Classification accuracy for *Experiment 2 (Part 1)* using the *fruit puree dataset*

Stimulus number <sup>a</sup>	Classification accuracy (%)	
	AM	FM
Stimulus 11 (A65, B2; A70)	61.90	95.24
Stimulus 12 (A79, B10; B13)	66.67	100.00
Stimulus 13 (A128, B13; B27)	47.62	95.24
Stimulus 14 (A518, B2; A529)	52.38	80.95
Stimulus 15 (A204, B36; B49)	61.90	85.71
Stimulus 16 (A524, B64; A529)	71.43	100.00
Stimulus 17 (A841, B49; A70)	28.57	85.71
Stimulus 18 (A857, B56; A743)	76.19	66.67
Stimulus 19 (A204, B49; B50)	57.14	76.19
Stimulus 20 (A547, B58; B60)	52.38	100.00
Mean (%)	57.62	88.57
Standard deviation (%)	13.55	11.49
Median (%)	59.52	90.48

<sup>a</sup> Stimulus number (Class A sample numbers, Class B sample numbers, Reference (X) sample numbers)



**Fig. 12** Boxplots of the sound synthesis methods using the *coffee* dataset and the *fruit puree* dataset

independent-samples *t*-test thus showed a significant difference in both dimensionality reduction methods.

A two-way between-groups analysis of variance (ANOVA) was conducted to explore the impact of different datasets and different dimensionality reduction methods on classification accuracy. There was no statistically significant main effect for the datasets used,  $F(1, 36) = 0.002$ ,  $p = 0.968$  and  $\eta_p^2 = 0.000$ . The dimensionality reduction method has,  $F(1, 36) = 305.809$ ,  $p < 0.001$  and  $\eta_p^2 = 0.895$ . This indicates clearly that the method used to reduce the dimensionality has the dominant impact in determining the accuracy of the classification. Overall, we conclude that the classification accuracy was improved by using the PCA method, regardless of what type of dataset was used.

A paired-samples *t*-test was conducted, which indicated a significant difference exists between the mean values in the AM sound synthesis ( $M = 57.62$ ,  $SD = 13.55$ ) and in the FM sound synthesis ( $M = 88.57$ ,  $SD = 11.49$ );  $t(9) = 5.28$ ,  $p = 0.001$ , which latter is  $< 0.05$ . The paired-samples *t*-test thus shows a significant difference in both sound synthesis methods.

A two-way between-groups analysis of variance (ANOVA) was conducted to explore the impact of different datasets and different sound synthesis methods on classification accuracy. There was no statistically significant main effect for the datasets used,  $F(1, 36) = 0.071$ ,  $p = 0.791$  and  $\eta_p^2 = 0.002$ . The sound synthesis methods have,  $F(1, 36) = 108.537$ ,  $p < 0.001$  and  $\eta_p^2 = 0.751$ . This indicates clearly that the method used to synthesise the sound has the dominant impact in determining the accuracy of the classification. Overall, we conclude that the classification accuracy was significantly improved in the FM synthesis, regardless of what type of dataset was used.

#### 4.3.4 Experiment 2 (Part 2): Fruit Puree Dataset

The results show that the mean accuracy obtained using PCA 88.57% whereas the mean accuracy obtained using the LDA is 42.50% as shown in Table 6 and Fig. 13. This strongly suggests that the classification accuracy for the PCA is once more significantly higher than for the LDA using the *fruit puree* dataset.

An independent-samples *t*-test was conducted, which indicated a significant difference between the mean values using PCA ( $M = 88.57\%$ ,  $SD = 11.49\%$ ) and LDA ( $M = 42.50\%$ ,  $SD = 10.36\%$ ); with  $t(18) = 9.418$ ,  $p < 0.001$ . The independent-samples *t*-test thus showed a significant difference in both dimensionality reduction methods.

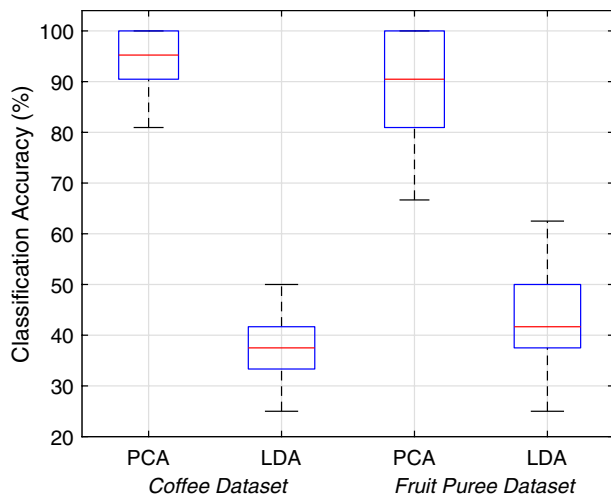
A two-way between-groups analysis of variance (ANOVA) was conducted to explore the impact of different datasets and different dimensionality reduction methods on classification accuracy. There was no statistically significant main effect for the datasets used,  $F(1, 36) = 0.002$ ,  $p = 0.968$  and  $\eta_p^2 = 0.000$ . The dimensionality reduction method has,  $F(1, 36) = 305.809$ ,  $p < 0.001$  and  $\eta_p^2 = 0.895$ . This indicates clearly that the method used to reduce the dimensionality has the dominant impact in determining the accuracy of the classification. Overall, we conclude that the classification accuracy was improved by using the PCA method, regardless of what type of dataset was used.

**Table 6** Classification accuracy for *Experiment 2 (Part 2)* using the *fruit puree* dataset

Stimulus number <sup>a</sup>	Classification accuracy (%)	
	PCA	LDA
Stimulus 1 (A65, B2; A70)	95.24	50.00
Stimulus 2 (A79, B10; B13)	100.00	62.50
Stimulus 3 (A128, B13; B27)	95.24	41.67
Stimulus 4 (A518, B2; A529)	80.95	37.50
Stimulus 5 (A204, B36; B49)	85.71	45.83
Stimulus 6 (A524, B64; A529)	100.00	33.33
Stimulus 7 (A841, B49; A70)	85.71	25.00
Stimulus 8 (A857, B56; A743)	66.67	37.50
Stimulus 9 (A204, B49; B50)	76.19	41.67
Stimulus 10 (A547, B58; B60)	100.00	50.00
Mean (%)	88.57	42.50
Standard deviation (%)	11.49	10.36
Median (%)	90.48	41.67

<sup>a</sup>Stimulus number (Class A sample numbers, Class B sample numbers, Reference (X) sample numbers)





**Fig. 13** Boxplots of the dimensionality reduction methods using the *coffee dataset* and the *fruit puree dataset*

## 5 Discussion

In this study, both AM synthesis and FM synthesis were used as sound synthesis methods in our sonification algorithm. Both AM and FM synthesis work in the same way as a parameter mapping technique. However, the difference between the synthesis methods is how the carrier wave is modulated or altered. This study suggests clearly that the FM synthesis method which adjusts the timbres of the sound, and leads to greater accuracy in the use of a sonification techniques. Thus, the FM synthesis method by using sound quality and timbre, has a significant advantage over the AM synthesis method.

Using FM synthesis in our sonification algorithm as sound synthesis, we then compare dimensionality reduction methods of PCA and LDA. The results suggest that the PCA method is more effective when being used in sonification technique for discriminating between two classes of food data. In general, an algorithm based on LDA tends to outperform compared to those based on PCA mainly for multi-class classification because the class labels are indicated and are taken into consideration Navarrete and Ruiz-del-Solar (2002); Belhumeur et al. (1997). In this study, we show that this is not always the case. One of the studies by Martínez and Kak (2001) asserts that PCA tends to perform better than LDA in classification accuracies for image recognition, provided when the number of the samples per class is relatively small. Traditionally, LDA has a problem called the small sample size problem (SSS) which occurs when the dimensions are much greater than the number of samples in the data matrix Howland and Park (2004). Thus, LDA fails to find a lower dimensional space and has difficulty in dealing with small sample data with high dimensionality. Some data sizes used by some researchers were not atypical

when PCA outperforms LDA. In a small sample size situation, the between-class scatter matrix ( $S_{between}$ ) or the within-class scatter matrix ( $S_{within}$ ) might be singular and LDA may encounter computational difficulty. This could explain why the accuracy of *Experiment 2 (Part 2)* using LDA is slightly higher than accuracy of classification in *Experiment 1 (Part 2)* because the dataset used in *Experiment 2* is the *fruit puree dataset* which has a higher dimensionality (sample size) compared to *Experiment 1* using the *coffee dataset*. However, overall the accuracy of classification using LDA is lower than PCA in both experiments. In addition, the classes of both datasets are not linearly separable in the original feature space. LDA fails to find the projection defining LDA space when the discriminatory information does not lie in the means of classes. Although the LDA method can effectively reduce the dimension of spectral data, it is less capable of dealing with classes which have non-linear decision boundaries because the LDA method requires the class separators to be hyperplanes. Due to the singularity in the within-class scatter matrix  $S_{within}$ , LDA is not an appropriate choice to be used in our sonification algorithm for dimensionality reduction.

Whilst this shows that participants were able to successfully classify samples using the FM synthesis method in our sonification algorithm, there is also a high variance which needs further investigation. Furthermore, the subjective nature of this approach may contribute to the performance of the test. Subjectivity may be due to age, ear sensitivity, or perhaps emotional impact: all these can affect the performance of each subject. A larger study or range of subjects can be considered in order to improve the classification results. Features of a specific demographic can be matched accordingly to a future study that allows the users to provide an auditory feedback approach.

## 6 Conclusion

In this paper, we have proposed a spectroscopic analysis model for the purpose of discriminating between the two classes, which uses sonification to discriminate between classes in two data sets. By turning the spectroscopy data into audio signals instead, it is easier to distinguish between the two classes of data and the process is much quicker. This study thus converts the spectroscopy data into sounds, rather than the usual visual data, such as graphical plots. The combination of spectral acquisition, spectral pre-processing, feature extraction, dimensionality reduction (PCA) and sound synthesis processing (with timbre parameters selected from the features in the absorption spectrum), demonstrate a high accuracy in classification using FM synthesis. The results of the listening test demonstrate the efficiency and accuracy of our approach. It

will be necessary to investigate the efficiency of choosing different features in the absorption spectrum to map to the timbre parameters of a synthesiser. The current listening test is not large enough to evaluate the whole model, which thus requires more detailed analysis. The proposed model can be further implemented and evaluated in the context of real-time food sample analysis.

## 7 Resources

The list of online versions of the listening test for the current study and with the same set of samples for each experiment as that used by participants can be accessed from the following link:

- (i) <http://dmtlab.bcu.ac.uk/hsein/waet/test.html?url=tests/sonificationlisteningtest2.xml>  
for Experiment 1 (Part 1)
- (ii) <http://dmtlab.bcu.ac.uk/hsein/waet/test.html?url=tests/sonificationlisteningtest3.xml>  
for Experiment 1 (Part 2)
- (iii) <http://dmtlab.bcu.ac.uk/hsein/waet/test.html?url=tests/sonificationlisteningtest5.xml>  
for Experiment 2 (Part 1)
- (iv) <http://dmtlab.bcu.ac.uk/hsein/waet/test.html?url=tests/sonificationlisteningtest6.xml>  
for Experiment 2 (Part 2)

**Acknowledgements** We would like to express our sincere gratitude to all the participants who took part in the experiment. We would also like to thank the referees whose helpful comments have significantly improved the clarity of this paper.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interests.

**Ethical approval** This article does not contain any studies with human participants performed by any of the authors.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ahmad, A., Adie, S. G., Wang, M., & Boppart, S. A. (2010). Sonification of optical coherence tomography data and images. *Optics Express*, 18(10), 9934–9944. <https://doi.org/10.1364/OE.18.009934>.
- Beghi, R., Giovanelli, G., Malegori, C., Giovenzana, V., & Guidetti, R. (2014). Testing of a VIS-NIR system for the monitoring of long-term apple storage. *Food and Bioprocess Technology*, 7(7), 2134–2143. <https://doi.org/10.1007/s11947-014-1294-x>.
- Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7), 711–720. <https://doi.org/10.1109/34.598228>.
- Briandet, R., Kemsley, E. K., & Wilson, R. H. (1996). Discrimination of Arabica and Robusta in instant coffee by Fourier transform infrared spectroscopy and chemometrics. *Journal of Agricultural and Food Chemistry*, 44(1), 170–174. <https://doi.org/10.1021/jf950305a>.
- Cassidy, R. J., Berger, J., Lee, K., Maggioni, M., & Coifman, R. R. (2004). Auditory display of hyperspectral colon tissue images using vocal synthesis models. In *Proceedings of ICAD 04-10th meeting of the international conference on auditory display*, Georgia Institute of Technology.
- Cassidy, R. J., Berger, J., Lee, K., Maggioni, M., & Coifman, R. R. (2004). Analysis of hyperspectral colon tissue images using vocal synthesis models. In *Conference record of the thirty-eighth asilomar conference on signals, systems and computers*, 2004, 2, 1611–1615, IEEE. <https://doi.org/10.1109/ACSSC.2004.1399429>.
- Childs, E., & Pulkki, V. (2003). Using multi-channel spatialization in sonification: A case study with meteorological data. In *Proceedings of the 2003 international conference on auditory display*, Georgia Institute of Technology.
- Clark, D. (1982). High-resolution subjective testing using a double-blind comparator. *Journal of the Audio Engineering Society*, 30(5), 330–338.
- de Campo A., Frauenberger c., & Höldrich R. (2004). Designing a generalized sonification environment. In *Proceedings of ICAD 04-tenth meeting of the international conference on auditory display*. Georgia Institute of Technology.
- de Campo, A., Höldrich, R., Sengl, B., Melde, T., Plessas, W., & Frauenberger, C. (2005). Sonification of quantum spectra. In *Proceedings of ICAD 05-eleventh meeting of the international conference on auditory display*. Georgia Institute of Technology.
- Defernez, M., & Wilson, R. H. (1995). Mid-infrared spectroscopy and chemometrics for determining the type of fruit used in jam. *Journal of the Science of Food and Agriculture*, 67(4), 461–467. <https://doi.org/10.1002/jsfa.2740670407>.
- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2), 179–188. <https://doi.org/10.1111/j.1469-1809.1936.tb02137.x>.
- Giovenzana, V., Beghi, R., Malegori, C., Civelli, R., & Guidetti, R. (2014). Wavelength selection with a view to a simplified handheld optical system to estimate grape ripeness. *American Journal of Enology and Viticulture*, 65(1), 117–123. <https://doi.org/10.5344/ajev.2013.13024>.
- Grond, F., & Berger, J. (2011). Parameter mapping sonification. In T. Hermann, A. Hunt, & J. G. Neuhoﬀ (Eds.), *The sonification handbook* (pp. 363–398). Berlin: Logos.
- Henrique, P. A., & Ruz, G. A. (2019). Noise reduction for near-infrared spectroscopy data using extreme learning machines. *Engineering Applications of Artificial Intelligence*, 79, 13–22. <https://doi.org/10.1016/j.engappai.2018.12.005>.
- Holland, J. K., Kemsley, E. K., & Wilson, R. H. (1998). Use of Fourier transform infrared spectroscopy and partial least squares

- regression for the detection of adulteration of strawberry purees. *Journal of the Science of Food and Agriculture*, 76(2), 263–269. [https://doi.org/10.1002/\(SICI\)1097-0010\(199802\)76:2<263::AID-JSFA943>3.0.CO;2-F](https://doi.org/10.1002/(SICI)1097-0010(199802)76:2<263::AID-JSFA943>3.0.CO;2-F).
- Howland, P., & Park, H. (2004). Generalizing discriminant analysis using the generalized singular value decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(8), 995–1006. <https://doi.org/10.1109/TPAMI.2004.46>.
- Jillings, N., Moffat, D., De Man, B., Reiss, J. D., & Stables, R. (2016). Web Audio Evaluation Tool: A framework for subjective assessment of audio. Web Audio Conference, Atlanta, USA.
- Khosravy, M., Gupta, N., Patel, N., Dey, N., Nitta, N., & Babaguchi, N. (2020). Probabilistic Stone's Blind Source Separation with application to channel estimation and multi-node identification in MIMO IoT green communication and multimedia systems. *Computer Communications*, 157, 423–433. <https://doi.org/10.1016/j.comcom.2020.04.042>.
- Kleiman-Weiner, M., & Berger, J. (2006). The sound of one arm swinging: A model for multidimensional auditory display of physical motion. In *Proceedings of the 12th international conference on auditory display*, Georgia Institute of Technology.
- Krzanowski, W. J., Jonathan, P., McCarthy, W. V., & Thomas, M. R. (1995). Discriminant analysis with singular covariance matrices: Methods and applications to spectroscopic data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 44(1), 101–115. <https://doi.org/10.2307/2986198>.
- Llabjani, V., Trevisan, J., Jones, K. C., Shore, R. F., & Martin, F. L. (2011). Derivation by infrared spectroscopy with multivariate analysis of bimodal contaminant-induced dose-response effects in MCF-7 cells. *Environmental Science & Technology*, 45(14), 6129–6135. <https://doi.org/10.1021/es200383a>.
- Martínez, A. M., & Kak, A. (2001). C, PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2), 228–233. <https://doi.org/10.1109/34.908974>.
- Navarrete, P., & Ruiz-del-Solar, J. (2002). Analysis and comparison of eigenspace-based face recognition approaches. *International Journal of Pattern Recognition and Artificial Intelligence*, 16(07), 817–830. <https://doi.org/10.1142/S0218001402002003>.
- Oliveri, P., Di Egidio, V., Woodcock, T., & Downey, G. (2011). Application of class-modelling techniques to near infrared data for food authentication purposes. *Food Chemistry*, 125(4), 1450–1456. <https://doi.org/10.1016/j.foodchem.2010.10.047>.
- O'Neill, C., & Ng, K. (2008). Hearing images: Interactive sonification interface for images. In *2008 international conference on automated solutions for cross media content and multi-channel distribution*, pp. 25–31, IEEE. <https://doi.org/10.1109/AXMEDIS.2008.42>.
- Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559–572. <https://doi.org/10.1080/14786440109462720>.
- Quadram Institute Bioscience, “Analytical techniques for food authentication,” (2007). Retrieved May 18, 2018 from <https://csr.quadram.ac.uk>.
- Rinnan, Å., Van Den Berg, F., & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*, 28(10), 1201–1222. <https://doi.org/10.1016/j.trac.2009.07.007>.
- Savitzky, A., & Golay, M. J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8), 1627–1639. <https://doi.org/10.1021/ac60214a047>.
- Sedaaghi, M. H., Daj, R., & Khosravi, M. (2001). Mediated morphological filters. In *Proceedings 2001 international conference on image processing (Cat. No. 01CH37205)*, 3, 692–695. <https://doi.org/10.1109/ICIP.2001.958213>.
- Stone, J. V. (2004). *Independent component analysis: A tutorial introduction*. New York: MIT.
- Stone, J. V. (2006). Blind source separation using temporal predictability. *Neural Computation*, 13(7), 1559–1574. <https://doi.org/10.1162/089976601750265009>.
- Strong, R., Martin, F. L., Jones, K. C., Shore, R. F., & Halsall, C. J. (2017). Subtle effects of environmental stress observed in the early life stages of the Common frog, *Rana temporaria*. *Scientific Reports*, 7(1), 1–13. <https://doi.org/10.1038/srep44438>.
- Sturm, B. L. (2002). Surf music: Sonification of ocean buoy spectral data. In *Proceedings of the 2002 international conference on auditory display*, Georgia Institute of Technology.
- Uslu, F. S., Binol, H., & Bal, A. (2016). Food inspection using hyperspectral imaging and SVDD. Sensing for Agriculture and Food Quality and Safety VIII. *International Society for Optics and Photonics*, 9864, 98640N. <https://doi.org/10.1117/12.2223938>.
- Vicinanza, D., Stables, R., Clemens, G., & Baker, M. (2014). Assisted differentiated stem cell classification in infrared spectroscopy using auditory feedback. In *The 20th international conference on auditory display (ICAD-2014)*, Georgia Institute of Technology.
- Zhang, Z. Y., Gui, D. D., Sha, M., Liu, J., & Wang, H. Y. (2019). Raman chemical feature extraction for quality control of dairy products. *Journal of Dairy Science*, 102(1), 68–76. <https://doi.org/10.3168/jds.2018-14569>.
- Zhongzhi, H. (2019). *Computer vision-based agriculture engineering*. New York: CRC.
- Zwicker, U. T. (1984). Auditory recognition of diotic and dichotic vowel pairs. *Speech Communication*, 3(4), 265–277. [https://doi.org/10.1016/0167-6393\(84\)90023-2](https://doi.org/10.1016/0167-6393(84)90023-2).

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.